

## SYSTEM AND METHOD FOR MINING A USER'S ELECTRONIC MAIL MESSAGES TO DETERMINE THE USER'S AFFINITIES

5

### CROSS REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Patent Application No. \_\_\_\_\_ (attorney docket no. 23452-500-301) entitled KNOWLEDGE SERVER, filed on January 14, 2002, the contents of which are incorporated by reference into this patent application.

This application is related to the following commonly owned U.S. patent applications, all of which are hereby incorporated by reference into the present application: (1) U.S. Patent Application No. 09/401,581, entitled METHOD AND SYSTEM FOR PROFILING USERS BASED ON THEIR RELATIONSHIP WITH CONTENT TOPICS, filed Sept. 22, 1999; (2) U.S. Patent Application No. \_\_\_\_\_ (attorney docket no. 23452-509) entitled METHOD AND SYSTEM FOR PROFILING USERS BASED ON THEIR RELATIONSHIP WITH CONTENT TOPICS, filed January 15, 2002; (3) U.S. Patent Application No. \_\_\_\_\_ (attorney docket no. 23452-507) entitled SYSTEM AND METHOD FOR PUBLISHING A PERSON'S AFFINITIES, filed January 15, 2002; (4) U.S. Patent Application No. \_\_\_\_\_ (attorney docket no. 23452-501) entitled SYSTEM AND METHOD FOR CALCULATING A USER AFFINITY, filed January 15, 2002; and (5) U.S. Patent Application No. \_\_\_\_\_ (attorney docket no. 23452-505) entitled SYSTEM AND METHOD FOR IMPLEMENTING A METRICS ENGINE FOR TRACKING RELATIONSHIPS OVER TIME, filed January 15, 2002.

### BACKGROUND OF THE INVENTION

#### 1. Field of the Invention

The present invention relates to the field of knowledge management, and, more specifically, to a system and method for mining a user's electronic mail messages to determine the user's affinities.

2. Discussion of the Background

When a person is attempting to accomplish a task, it is often useful for the person to obtain information from other people who have knowledge of the topics with which the task is concerned. To do so, the person must have a way to discover the people who have the information the person is seeking to obtain. One way of facilitating this discovery is to publish people's "affinities," which are simply links between people and categories or topics of information. Each affinity may include a value representing the strength of the relationship with the category--the higher the value, the greater the person's affinity for the topic.

It is possible that publishing a person's affinity (i.e., making the affinity known to others) would be inappropriate, either because the affinity is inaccurate or misleading, or because it reveals an accurate relationship with a topic that the person does not wish to make public. Therefore, it is important to provide ways for people to judge their proposed affinities accurately and to avoid affinity publication in such cases. Recognizing that policies concerning affinity publication may be affected by different cultures and laws, the solution to these problems must be flexible as well.

20 SUMMARY OF THE INVENTION

The present invention provides a system and method for mining a user's e-mail (i.e., examining the content of the user's e-mail) and for generating a list of concepts (also referred to as categories) based on the e-mail content. The generated category list is compared to a master category list and those categories included in the generated category list that are not included in the master category list are removed from the generated category list. For each category remaining in the generated category list, the system and method calculates an affinity value, which represents the strength of the user's relationship to the category. The affinity (i.e., the category plus the affinity value) may be submitted to an affinity publisher module that uses an affinity publication policy in determining whether or not to publish the affinity.

In one aspect, a method according to the present invention for mining e-mails to determine a user's affinities includes the following steps: accessing an e-mail system and retrieving from the system the e-mails sent to and from the user; extracting 5 keywords from the retrieved e-mails; generating a list of categories based on the extracted keywords; accessing a master category list; filtering the generated category list by removing from the generated list those categories that are not included in the master category list; and for each category remaining in the generated category list, calculating an affinity value, associating the affinity value with the category, and submitting the category and the affinity value to the affinity publisher module.

The above and other features and advantages of the present invention, as well as the structure and operation of various embodiments of the present invention, are described in detail below with reference to the accompanying drawings.

#### BRIEF DESCRIPTION OF THE DRAWINGS

The accompanying drawings, which are incorporated herein and form part of the specification, illustrate various embodiments of the present invention and, together with the description, further serve to explain the principles of the invention and to enable a person skilled in the pertinent art to make and use the invention. In the drawings, like 20 reference numbers indicate identical or functionally similar elements. Additionally, the left-most digit(s) of a reference number identifies the drawing in which the reference number first appears.

FIG. 1 is a functional block diagram of a system according to one embodiment 25 of the present invention.

FIG. 2 is a flow chart illustrating a process, according to one embodiment, performed by affinity publisher module.

FIG. 3 is a flow chart illustrating a process, according to one embodiment, for publishing a designated affinity.

5 FIG. 4 is a flow chart illustrating a process, according to one embodiment, for enabling a user to declare and publish an affinity.

FIG. 5 is a flow chart illustrating a process, according to one embodiment, for mining electronic mail (e-mail).

10 FIG. 6 is a flow chart illustrating a process, according to one embodiment, for creating a master category list.

#### DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

While the present invention may be embodied in many different forms, there is described herein in detail an illustrative embodiment with the understanding that the present disclosure is to be considered as an example of the principles of the invention and is not intended to limit the invention to the illustrated embodiment.

20 FIG. 1 is a functional block diagram of a system 100 according to one embodiment of the present invention. System 100 includes a computer system 150 for executing an affinity publisher software module 102 and an affinity discovery software module 104, an affinity publication policy 106, and a storage system 108 that stores a plurality of user profiles 110 and a plurality of category profiles 166, wherein each user profile 110 is a set of information that is associated with a particular user (e.g., profile 110(b) is associated with a user 101), and wherein each category profile 166 is a set of information associated with a particular category. Storage system 108 includes one or more storage devices so that user profiles 110 and category profiles 166 need not be stored on the same storage device. A profile can be a single computer file, one or more computer files, one or more records in a database, etc. Computer system 150 includes 25 one or more computers (not shown).

30

Affinity discovery module 104 functions to monitor the activities of user 101 to determine the subject matters (i.e., categories) for which user 101 appears to have an affinity, determines the strength of the affinity for each determined category, and assigns an affinity value to the determined affinity. As an example, affinity discovery module 104 may be operable to access an electronic mail (e-mail) system 187 to examine the e-mails sent to and from user 101 and may be operable to access a document repository 189 to examine the documents authored or viewed by user 101.

5 For example, if user 101 has recently authored and viewed several documents associated with the category of “computer security,” then affinity discovery module 104 will know this because it monitors user 101’s document activity. Consequently, affinity discovery module 104 will determine that user 101 appears to have an affinity for “computer security” based on user 101’s document activity. Additionally, affinity discovery module 104 will assign an affinity value to the affinity. The affinity value represents the strength of user 101’s affinity for the category.

10  
15 After affinity discovery module 104 determines that user 101 appears to have an affinity for a particular category and assigns an affinity value to the affinity, module 104 submits the “affinity” to affinity publisher module 102. That is, module 104 submits the name of the category and the calculated affinity value to module 102.

20 Upon receiving a submitted affinity, affinity publisher module 102 applies an affinity publication policy 106 to determine whether it should publish user 101’s apparent affinity for the particular category. Affinity publication policy 106 includes rules and other information that govern the publication of affinities. In one embodiment, publication policy 106 can only be created and modified by an affinity administrator 103. In other embodiments, affinity administrator 103 as well as other users can create and/or modify the affinity publication policy.

25

30 Affinity publication policy 106 preferably includes some or all of the following information: an affinity threshold value, an indication as to whether publisher module

102 must get permission from a user prior to publishing the user's affinities, an auto-response grace period, a setting for an auto-publish flag, and other information. Other information and other rules can be included in publication policy 106. The ability of administrator 103 to create an affinity publication policy creates a unique advantage because this features allows system 100 to be flexible and, thus, easily adapt to different cultures and laws regarding publication of private information.

If, based on affinity publication policy 106, module 102 determines that it should publish user 101's apparent affinity for the particular category, then, in one embodiment, module 102 updates one or both of the user profile 110 associated with user 101 (e.g., user profile 110(b)) and the category profile 166 associated with the particular category, so that the update profile indicates that user 101 has an affinity for the particular category. The user profile 110 and/or category profile 166 is/are also updated to indicate the affinity value assigned to the affinity.

Profiles 110 and 166 may be searched by third parties or search engines. In this way, after affinity publisher module 102 publishes user 101's affinity for the particular subject matter, a third person or a search engine or other system is able to determine that user 101 has an affinity for the particular category simply by examining profiles 110 and/or 166. In this way, a person who seeks to discover individuals who are likely to have knowledge and/or expertise about a certain topic can easily do so simply by searching profiles 110/166.

In one embodiment, system 100 includes a single affinity publication policy 106 (also referred to as "default affinity publication policy 106") that applies to all users whose activities are being monitored. In another embodiment, a user whose activities are being monitored may have his or her own affinity publication policy which overrides the default affinity publication policy. That is, when a user has his or her own affinity publication policy, affinity publisher module 102 uses that affinity publication policy instead of the default affinity publication policy in determining whether or not to

publish an affinity for the user.

FIG. 2 is a flow chart illustrating a process 200 performed by affinity publisher module 102 after discovery module 104 determines that user 101 appears to have an affinity for a particular category, assigns an affinity value for the apparent affinity, and submits the affinity to module 102. Process 200 begins in step 202, where module 102 determines whether user 101 has his or her own affinity publication policy. If user 101 has his or her own affinity publication policy, module 102 selects that affinity publication policy (step 204), otherwise, module 102 selects default affinity publication policy 106 (step 206). Next (step 208), module 102 determines the selected policy's affinity threshold. Next (step 210), module 102 determines whether the affinity value assigned by discovery module 104 exceeds the determined affinity threshold. If the assigned affinity value does not exceed the affinity threshold, the process ends, otherwise the process continues in step 212.

In step 212, module 102 determines whether the publication policy indicates that module 102 must get permission from user 101 prior to publishing user 101's affinities. If the publication policy indicates that module 102 must get permission from user 101 prior to publishing user 101's affinities, then control passes to step 214, otherwise control passes to step 224.

In step 214, module 102 notifies user 101 of user 101's apparent affinity for the particular category and requests permission from user 101 to publish the affinity. In one embodiment, as described above, a category profile, such as profile 166(b) is associated with the particular category. Category profile 166(b) may include: the names of all of the people that have a published affinity for the particular category, the names of the documents (if any) that are linked with or associated with the particular category, and information concerning the relationship between the particular category and other categories. In this embodiment, module 102 may send to user 101 the information included in category profile 166(b) along with the affinity notification because user 101

may find the information included in category profile 166(b) useful when determining the accuracy of the affinity and whether or not to approve publication of the affinity. In one embodiment, the affinity notification sent to user 101 includes not only the name of a category and an affinity value associated with the category, but also one or more keywords that are associated with the category. This additional information gives user 101 a better context for determining whether or not he or she wants to have the affinity published.

Next (step 216), module 102 determines the auto-response grace period for the selected affinity publication policy and sets a timer to expire when an amount of time equal to the grace period has elapsed. Next (step 218), module 102 waits for a response from user 101 or for the timer to expire. If a response is received before the timer expires, control passes to step 220, otherwise control passes to step 222.

In step 220, module 102 determines whether the response indicates that user 101 has approved the publication of the affinity. If the response indicates that user 101 has approved the publication of the affinity, control passes to step 224, otherwise the process ends.

In step 222, module 102 determines whether the selected affinity publication policy's auto-publish flag is set to TRUE. If it is, control passes to step 224, otherwise control passes to step 223, where module 102 notifies user 101 that the affinity will not be published because the grace period has expired. The process ends after step 223.

In step 224, module 102 publishes the affinity. In one embodiment, module 102 publishes the affinity by updating profile 110(b), which is associated with user 101, such that profile 110(b) indicates that user 101 has an affinity for the particular category. Advantageously, profile 110(b) may also be updated to indicate the strength of the affinity. That is, for example, the affinity value assigned to the affinity can be included in profile 110(b) along with the information that indicates user 101 has an

affinity for the category. After the affinity is published, module 102 may notify user 101 that the affinity was published (step 225). Preferably, in addition to (or instead of) updating profile 110(b), module 102 updates the category profile 166 that is associated with the particular category so that the category profile indicates that user 101 has an affinity for the particular category.

5

FIG. 3 is a flow chart illustrating a process 300 for publishing a designated affinity for user 101. A designated affinity for user 101 is an affinity assigned to user 101 by a third-party, such as user 101's manager, who may wish to assign an affinity to user 101.

10

Process 300 begins in step 302, where user 105 selects a category, submits the category to module 102, and requests module 102 to update user 101's profile (i.e., profile 110(b)) to indicate that user 101 has an affinity for the submitted category. In step 306, module 102 determines whether user 105 is authorized to designate an affinity for user 101. If user 105 is not so authorized, process 300 ends, otherwise control passes to step 310. In one embodiment, module 102 determines whether user 105 is authorized to designate affinities for user 101 by examining an affinity designator list 190. Preferably, administrator 103 controls the list and authorizes a user (such as user 105) to designate affinities for another user (such as user 101) by adding an entry to list 190 that indicates that the user has permission to designate affinities for the other user.

20

In step 310, module 102 either selects an affinity value or requests user 105 to input an affinity value. In step 312, module 102 determines whether user 101 has his or her own affinity publication policy, and, if user 101 has his or her own affinity publication policy, selects that affinity publication policy, otherwise, selects default affinity publication policy 106.

25

In step 318, module 102 determines whether the publication policy indicates that module 102 must get permission from user 101 prior to publishing the designated

30

affinity. If the publication policy indicates that module 102 must get permission from user 101 prior to publishing the designated affinity, then control passes to step 320, otherwise control passes to step 330.

5 In step 320, module 102 notifies user 101 of the proposed designated affinity and requests permission from user 101 to publish the affinity. In step 322, module 102 determines the selected affinity publication policy's auto-response grace period and sets a timer to expire when an amount of time equal to the grace period has elapsed. In step 324, module 102 waits for a response from user 101 or for the timer to expire. If a response is received before the timer expires, control passes to step 326, otherwise control passes to step 328.

10  
15 In step 326, module 102 determines whether the response indicates that user 101 has approved the publication of the designated affinity. If the response indicates that user 101 has approved the publication of the designated affinity, control passes to step 330, otherwise the process ends.

20 In step 328, module 102 determines whether the selected affinity publication policy's auto-publish flag is set to TRUE. If it is, control passes to step 330, otherwise control passes to step 329, where module 102 notifies user 101 that the affinity will not be published because the grace period has expired. The process ends after step 329.

25 In step 330, module 102 publishes the designated affinity. In one embodiment, module 102 publishes the designated affinity by updating profile 110(b), which associated with user 101, such that profile 110(b) indicates that user 101 has an affinity for the submitted category. Advantageously, profile 110(b) may also be updated to indicate the strength of the affinity. That is, for example, the affinity value obtained in step 310 can be included in profile 110(b) along with the information that indicates user 101 has an affinity for the category. After the affinity is published, user 101 may be notified that the affinity was published (step 331). Preferably, in addition to (or instead

of) updating profile 110(b), module 102 updates the category profile 166 that is associated with the particular category so that the category profile indicates that user 101 has an affinity for the particular category.

5 In addition to publishing derived affinities (that is, affinities determined by affinity discovery module 104) and designated affinities, module 102 can be configured to allow a user to declare his or her own affinities. FIG. 4 is a flow chart illustrating a process for enabling user 101 to declare and publish an affinity. Process 400 begins in step 402, where user 101 selects a category. In step 404, user submits the selected category to module 102. In step 406, module 102 either selects an affinity value or requests user 101 to submit an affinity value. In step 408, module 102 publishes the designated affinity.

FIG. 5 is a flow chart illustrating a process 500, which may be performed by affinity discovery module 104, for mining electronic mail (e-mail) for the purpose of determining a user's affinities. Process 500 begins in step 502, where module 104 accesses e-mail system 187 and retrieves the e-mails sent to and from the user. Next (step 504), module 104 extracts keywords from the retrieved e-mails. Next (step 506), module 104 generates a list of categories (or concepts) based on the extracted keywords. Next (step 508), module 102 access a master category list 168. Next (step 510), module 104 filters the category list generated in step 506 by removing from the list the categories that are not included in the master category list. Next (step 512), for each category remaining in the generated category list, module 104 calculates an affinity value, associates the affinity value with the category, and submits the category and the affinity value to affinity publisher module 102, which then performs process 200.

25 The feature of filtering the category list generated in step 506 based on the master category list provides a mechanism for protecting the user's privacy. It protects the user's privacy by ensuring that only the user's affinity for categories included in the master category list have a chance of being published. In other words, there is no

chance that affinity publisher module 102 will publish the user's affinity for a category that is not on the master category list. In this way, system 100 provides privacy protection.

5 In one embodiment, when module 104 is mining the e-mails received by and/or sent from a particular user, module 104 uses keywords generated from the content of one or more of those e-mails to determine affinities for other users who also received or sent those e-mails. For example, if 15 of the e-mails received by user A were also received by or sent from user B, then when module 104 is mining user A's e-mails module 104 can use these 15 e-mails to discover affinities for user B. In this way, module 104 can determine affinities for user B based on e-mail content even if user B has not given module 104 permission to mine his or her e-mails.

10 FIG. 6 is a flow chart illustrating a process 600, according to one embodiment, for creating master category list 168. Process 600 begins in step 602, where a set of documents from one or more document repositories (such as repository 189) are accessed. In one embodiment, the set of documents may be selected by administrator 103, but in other embodiments the set of documents are selected according to other criteria, such as the author and/or type of document. Next (step 604) keywords are extracted from the set of documents. Next, (step 606), a list of categories (or concepts) based on the extracted keywords is generated. Lastly (step 608), categories can be manually added to and/or deleted from the list as desired.

15 While the illustrated processes 200, 300, 400, 500 and 600 are described as a series of consecutive steps, none of these processes are limited to any particular order of the described steps. Additionally, it should be understood that the various illustrative embodiments of the present invention described above have been presented by way of example only, and not limitation. Thus, the breadth and scope of the present invention should not be limited by any of the above-described exemplary embodiments, but 20 should be defined only in accordance with the following claims and their equivalents.